Higher order derivatives and perturbation bounds for some functions of matrices

Priyanka Grover*

Indian Statistical Institute, Delhi Centre, 7, S.J.S. Sansanwal Marg, New Delhi-110016, India

July 27, 2010

Abstract

In this paper, we obtain three different expressions for all higher order derivatives of the permanent of a matrix and coefficients of its characteristic polynomial. Upper bound for the norms of the derivatives of the permanent is given. Norms of the derivatives for coefficients of the characteristic polynomial are evaluated exactly.

1 Introduction

The Jacobi formula for the derivative of the determinant of a matrix is well known. Recently, Bhatia and Jain have obtained analogous expressions for its higher order derivatives (See [7]). They derive three formulas each of which is a generalisation of the Jacobi formula.

In this note, we extend these results in two directions. First we obtain formulas for derivatives of all orders for the permanent function. Second, we obtain similar formulas for all the coefficients of the characteristic polynomial of a matrix. These formulas then lead to higher order perturbation bounds for the functions studied.

Let $A = (a_{ij})$ be an $n \times n$ complex matrix. Let $\phi : \mathbb{M}(n) \to \mathbb{C}$ be a differentiable map. For each $X \in \mathbb{M}(n)$,

$$\mathrm{D}\phi(A)(X) = \frac{d}{dt}\bigg|_{t=0} \phi(A + tX). \tag{1.1}$$

Then $\mathrm{D}\phi$ is a linear map of $\mathbb{M}(n)$ into $\mathscr{L}(\mathbb{M}(n);\mathbb{C})$, the space of all linear operators from $\mathbb{M}(n)$ into $\mathbb{C}$. The *second derivative* of ϕ at A is the derivative of $\mathrm{D}\phi$ at A and is denoted by $\mathrm{D}^2\phi(A)$. This is an element of $\mathscr{L}(\mathbb{M}(n);\mathscr{L}(\mathbb{M}(n);\mathbb{C}))$ which is identified with $\mathscr{L}_2(\mathbb{M}(n);\mathbb{C})$, the space of bilinear mappings of $\mathbb{M}(n) \times \mathbb{M}(n)$ into $\mathbb{C}$. Similarly, for any k, the k^{th}

Email address: pgrover8r@isid.ac.in

derivative of ϕ at A, denoted by $\mathrm{D}^k\phi(A)$, is an element of $\mathscr{L}_k(\mathbb{M}(n);\mathbb{C})$, the space of multilinear mappings of $\mathbb{M}(n)\times\cdots\times\mathbb{M}(n)$ into $\mathbb{C}$ (See [8]). For $X^1,\ldots,X^k\in\mathbb{M}(n)$,

$$\mathrm{D}^k\phi(A)(X^1,\ldots,X^k)=\left.\frac{\partial^k}{\partial t_1\cdots\partial t_k}\right|_{t_1=\cdots=t_k=0}\phi(A+t_1X^1+\cdots+t_kX^k). \tag{1.2}$$

Notations. Let

$$Q_{k,n}=\{(i_1,\ldots,i_k)|1\le i_1<\cdots<i_k\le n\}$$

and

$$G_{k,n}=\{(i_1,\ldots,i_k)|1\le i_1\le\cdots\le i_k\le n\}.$$

For $k>n$, $Q_{k,n}=\Phi$, by convention. Also, note here that for $k\le n, Q_{k,n}$ is a subset of $G_{k,n}$. Given two elements $\mathcal{I}$ and $\mathcal{J}$ of $G_{k,n}$, let $A[\mathcal{I}|\mathcal{J}]$ denote the $k\times k$ matrix whose (r,s)-entry is the (i_r,j_s)-entry of A. If $\mathcal{I},\mathcal{J}\in Q_{k,n}$, then $A[\mathcal{I}|\mathcal{J}]$ is a submatrix of A. The rest of the notations are kept same as those in [7]:

If $\mathcal{I},\mathcal{J}\in Q_{k,n}$, then we denote by $A(\mathcal{I}|\mathcal{J})$, the $(n-k)\times(n-k)$ submatrix obtained from A by deleting rows $\mathcal{I}$ and columns $\mathcal{J}$. The j^{th} column of the matrix X is denoted by $X_{[j]}$. Given $n\times n$ matrices $X^1,\ldots,X^k$ and $\mathcal{J}=(j_1,\ldots,j_k)\in Q_{k,n}$, we use $A(\mathcal{J};X^1,\ldots,X^k)$ to mean the matrix obtained from A by replacing the j_p^{th} column of A by the j_p^{th} column of X^p for $1\le p\le k$, and keeping the rest of the columns unchanged, that is, if $Z=A(\mathcal{J};X^1,\ldots,X^k)$, then $Z_{[j_p]}=X^p_{[j_p]}$ for $1\le p\le k$, and $Z_{[l]}=A_{[l]}$ if l does not occur in $\mathcal{J}$. Let σ be a permutation on k symbols, then by $Y^\sigma_{[\mathcal{J}]}$, we mean the matrix in which $Y^\sigma_{[j_p]}=X^{\sigma(p)}_{[j_p]}$ for $1\le p\le k$ and $Y^\sigma_{[l]}=0$ if l does not occur in $\mathcal{J}$.

2 Permanent

The *permanent* of A, written as $\operatorname{per}(A)$, or simply $\operatorname{per}A$, is defined by

$$\operatorname{per}A=\sum_\sigma a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}, \tag{2.1}$$

where the summation extends over all the permutations of 1,2,. . . ,n.

Let $\operatorname{per}:\mathbb{M}(n)\to\mathbb{C}$ be the map taking an $n\times n$ matrix to its permanent. This is a differentiable map. We denote by $\mathrm{D}\operatorname{per}A$, the derivative of per at A.

The famous *Jacobi formula* for determinant says that

$$\mathrm{D}\det A(X)=\operatorname{tr}(\operatorname{adj}(A)X), \tag{2.2}$$

where the symbol $\mathrm{adj}\,(A)$ stands for the *adjugate*(the *classical adjoint*) of A. The *permanental adjoint*, denoted by $\mathrm{padj}\,(A)$, is the $n \times n$ matrix whose (i,j)-entry is $\mathrm{per}\, A(i|j)$ (See [10], page 237). We obtain the following result similar to the *Jacobi formula* for determinant.

Theorem 2.1. *For each $X \in \mathbb{M}(n)$,*

$$\mathrm{D}\,\mathrm{per}\,(A)(X) = \mathrm{tr}(\mathrm{padj}(A)^T X). \tag{2.3}$$

Proof. For $1 \leq i \leq n$, let $A(j;X)$ be the matrix obtained from A by replacing the j^{th} column of A by the j^{th} column of X and keeping the rest of the columns unchanged. Then (2.3) can be restated as

$$\mathrm{D}\,\mathrm{per}\,(A)(X) = \sum_{j=1}^{n} \mathrm{per}\, A(j;X). \tag{2.4}$$

From (1.1), we note that $\mathrm{D}\,\mathrm{per}\, A(X)$ is the coefficient of t in the polynomial $\mathrm{per}\,(A+tX)$. Using the fact that per is a linear function of each of its columns, we immediately obtain (2.4).

□

The *Laplace expansion theorem* for permanents ([11], page 16) says that for any $\mathcal{I} \in Q_{k,n}$,

$$\mathrm{per}\, A = \sum_{\mathcal{J} \in Q_{k,n}} \mathrm{per}\, A[\mathcal{I}|\mathcal{J}]\, \mathrm{per}\, A(\mathcal{I}|\mathcal{J}). \tag{2.5}$$

In particular, for any $i, 1 \leq i \leq n$,

$$\mathrm{per}\, A = \sum_{j=1}^{n} a_{ij}\ \mathrm{per}\,(A(i|j)). \tag{2.6}$$

Using this, equation (2.4) can be rewritten as

$$\mathrm{D}\,\mathrm{per}\,(A)(X) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}\, \mathrm{per}\, A(i|j). \tag{2.7}$$

The following two theorems are analogues of theorems 1 and 2 of [7] and also generalisations of (2.4) and (2.7) respectively. Their proofs imitate the proofs in [7].

Theorem 2.2. *For $1 \leq k \leq n$,*

$$\mathrm{D}^k\, \mathrm{per}\, A(X^1, \ldots, X^k) = \sum_{\sigma \in S_k} \sum_{\mathcal{J} \in Q_{k,n}} \mathrm{per}\, A(\mathcal{J}; X^{\sigma(1)}, X^{\sigma(2)}, \ldots, X^{\sigma(k)}). \tag{2.8}$$

In particular,

$$\mathrm{D}^k\, \mathrm{per}\, A(X, \ldots, X) = k! \sum_{\mathcal{J} \in Q_{k,n}} \mathrm{per}\, A(\mathcal{J}; X, \ldots, X).$$

Proof. From (1.2), it follows that $\mathrm{D}^k \operatorname{per} A(X^1, \ldots, X^k)$ is the coefficient of $t_1 \cdots t_k$ in the expansion of $\operatorname{per}(A + t_1X^1 + \cdots + t_kX^k)$. Also using linearity of the per function in each of its columns, we obtain (2.8). ☐

Theorem 2.3. *For $1 \leq k \leq n$,*

$$\mathrm{D}^k \operatorname{per} A(X^1, \ldots, X^k) = \sum_{\sigma \in S_k} \sum_{\mathcal{I}, \mathcal{J} \in Q_{k,n}} \operatorname{per} A(\mathcal{I}|\mathcal{J}) \operatorname{per} Y^{\sigma}_{[\mathcal{J}]}[\mathcal{I}|\mathcal{J}]. \tag{2.9}$$

In particular,

$$\mathrm{D}^k \operatorname{per} A(X, \ldots, X) = k! \sum_{\mathcal{I}, \mathcal{J} \in Q_{k,n}} \operatorname{per} A(\mathcal{I}|\mathcal{J}) \operatorname{per} X[\mathcal{I}|\mathcal{J}].$$

Proof. For each $\mathcal{J} \in Q_{k,n}$, the Laplace expansion theorem gives

$$\operatorname{per} A(\mathcal{J}; X^{\sigma(1)}, \ldots, X^{\sigma(k)}) = \sum_{\mathcal{I} \in Q_{k,n}} \operatorname{per} A(\mathcal{I}|\mathcal{J}) \operatorname{per} Y^{\sigma}_{[\mathcal{J}]}[\mathcal{I}|\mathcal{J}].$$

Equation (2.9) is obtained by expanding each term in the summation of (2.8) in this way. ☐

We note here that

$$\mathrm{D}^n \operatorname{per} A(X, \ldots, X) = n! \operatorname{per} X \tag{2.10}$$

and

$$\mathrm{D}^k \operatorname{per} A(X, \ldots, X) = 0 \;\; \forall\, k > n. \tag{2.11}$$

We now describe a generalisation of (2.3). Let $\mathcal{H}$ be an n-dimensional Hilbert space. Let $\otimes^k\mathcal{H}$ be the k-fold tensor power of $\mathcal{H}$ and $\vee^k\mathcal{H}$ be the symmetric tensor power of $\mathcal{H}$. If $\{e_i\}$ is an orthonormal basis of $\mathcal{H}$, then for $\mathcal{I} = (i_1, \ldots, i_k) \in G_{k,n}$, we define $e_{\mathcal{I}} = e_{i_1} \vee \cdots \vee e_{i_k}$. If $\mathcal{I}$ consists of l distinct indices $i_1, \ldots, i_l$ with multiplicities $m_1, \ldots, m_l$ respectively, put $m(\mathcal{I}) = m_1! \cdots m_l!$. Note that if $\mathcal{I} \in Q_{k,n}$, then $m(\mathcal{I}) = 1$. An orthonormal basis of $\vee^k\mathcal{H}$ is $\{m(\mathcal{I})^{-1/2}e_{\mathcal{I}} : \mathcal{I} \in G_{k,n}\}$. It is conventional to order these multi-indices lexicographically. (See [4], Chapter 2.)

Let $\vee^k A$ denote the kth *symmetric tensor power* of A. With respect to the above mentioned basis, the $(\mathcal{I}, \mathcal{J})$-entry of $\vee^k A$ is $(m(\mathcal{I})m(\mathcal{J}))^{-1/2} \operatorname{per} A[\mathcal{I}|\mathcal{J}]$. Let P_k be the canonical projection of $\vee^k\mathcal{H}$ onto the subspace generated by $\{e_{\mathcal{I}} : \mathcal{I} \in Q_{k,n}\}$. If we vary $\mathcal{I}, \mathcal{J}$ in $Q_{k,n}$, we get the submatrix $P_k(\vee^k A)P_k$ of $\vee^k A$.

The matrix $\operatorname{padj}(A)^T$ can be identified with a submatrix of an operator on the space $\vee^{n-1}\mathcal{H}$. We call this operator $\tilde{\vee}^{n-1}A$. Then $\operatorname{padj}(A)^T = P_{n-1}(\tilde{\vee}^{n-1}A)P_{n-1}$, which is an $n \times n$ matrix. It is unitarily similar to the transpose of the submatrix $P_{n-1}(\vee^{n-1}A)P_{n-1}$ of $\vee^{n-1}A$. Similarly, the transpose of the matrix whose $(\mathcal{I}, \mathcal{J})$-entry is $\operatorname{per} A(\mathcal{I}|\mathcal{J})$ can be identified as a

submatrix of an operator on the space $\vee^{n-k}\mathcal{H}$. We call this operator $\tilde{\vee}^{n-k}A$. The $\binom{n}{k} \times \binom{n}{k}$ submatrix $P_{n-k}(\tilde{\vee}^{n-k}A)P_{n-k}$ of $\tilde{\vee}^{n-k}A$ is unitarily similar to the transpose of the submatrix $P_{n-k}(\vee^{n-k}A)P_{n-k}$ of $\vee^{n-k}A$.

Equation (2.3) can also be written as

$$\mathrm{D}\,\mathrm{per}\,A(X) = \mathrm{tr}\,(P_{n-1}(\tilde{\vee}^{n-1}A)P_{n-1})X. \tag{2.12}$$

Let $X^1, \ldots, X^k \in \mathscr{L}(\mathcal{H})$. Consider the following operator on $\otimes^k\mathcal{H}$:

$$\frac{1}{k!}\sum_{\sigma\in S_k} X^{\sigma(1)} \otimes X^{\sigma(2)} \otimes \cdots \otimes X^{\sigma(k)}. \tag{2.13}$$

It leaves the space $\vee^k\mathcal{H}$ invariant. We use the notation $X^1 \vee X^2 \vee \cdots \vee X^k$ for the restriction of this operator to the subspace $\vee^k\mathcal{H}$.

The generalisation of (2.12) is given as follows:

Theorem 2.4. *For* $1 \leq k \leq n$*,*

$$\mathrm{D}^k\,\mathrm{per}\,A(X^1, \ldots, X^k) = k!\,\mathrm{tr}\,[(P_{n-k}(\tilde{\vee}^{n-k}A)P_{n-k})(P_k(X^1 \vee \cdots \vee X^k)P_k)]. \tag{2.14}$$

In particular,

$$\mathrm{D}^k\,\mathrm{per}\,A(X, \ldots, X) = k!\,\mathrm{tr}\,[(P_{n-k}(\tilde{\vee}^{n-k}A)P_{n-k})(P_k(\vee^k X)P_k)].$$

Proof. Note that

$$Y^{\sigma}_{[\mathcal{J}]}[\mathcal{I}|\mathcal{J}] = (< e_{i_l}, X^{\sigma(m)}e_{j_m} >)_{1\leq l,m\leq k}$$

and $(\mathcal{I},\mathcal{J})$-element of $P_k(X^1 \vee \cdots \vee X^k)P_k$, where $\mathcal{I} = (i_1, \ldots, i_k)$ and $\mathcal{J} = (j_1, \ldots, j_k) \in Q_{k,n}$, is given by

$$\begin{aligned}
& < e_{\mathcal{I}}, (P_k(X^1 \vee \cdots \vee X^k)P_k)e_{\mathcal{J}} > \\
= \; & < e_{i_1} \vee \cdots \vee e_{i_k}, (X^1 \vee \cdots \vee X^k)(e_{j_1} \vee \cdots \vee e_{j_k}) > \\
= \; & \sum_{\sigma\in S_k} \frac{1}{k!}\,\mathrm{per}(< e_{i_l}, X^{\sigma(m)}e_{j_m} >)_{1\leq l,m\leq k}.
\end{aligned}$$

Also, $(\mathcal{J},\mathcal{I})$-element of $P_{n-k}(\tilde{\vee}^{n-k}A)P_{n-k}$ is $\mathrm{per}\,A(\mathcal{I}|\mathcal{J})$, by definition of $\tilde{\vee}^{n-k}A$. Using all this in (2.9), we obtain (2.14). $\square$

Bhatia and Dias da Silva (Theorem 1, [5]) have obtained norm of $\mathrm{D}K_\lambda(A)$ where K_λ is the restriction of the map $\otimes^k A$ to the symmetry class of tensors associated with λ and S_m. A particular case is theorem 1 in [1] which says that

$$\|\mathrm{D}\vee^k(A)\| = k\|A\|^{k-1} \quad \forall 1 \leq k \leq n. \tag{2.15}$$

It follows from (2.15) that

$$\|\mathrm{D}\,\mathrm{per}\,A\| \leq n\|A\|^{n-1}. \tag{2.16}$$

We extend this to the following:

Theorem 2.5. *Let A be an $n \times n$ matrix, we have*

$$\|\mathrm{D}^k \operatorname{per} A\| \leq k! \binom{n}{k} \|A\|^{n-k}. \tag{2.17}$$

Proof. To show this, we first note that if A is an $n \times n$ matrix, then

$$\|A\| \leq \|A\|_1 \leq n \, \|A\|. \tag{2.18}$$

Also, if $s_1(A) \geq \cdots \geq s_n(A)$ are the singular values of A, then the singular values of $\vee^k A$ are $s_{i_1} \cdots s_{i_k}$, where $(i_1, \ldots, i_n)$ vary over $G_{k,n}$. So,

$$\| \vee^k A\| = s_1(A)^k. \tag{2.19}$$

Now, by definition,

$$\|\mathrm{D}^k \operatorname{per} A\| = \sup_{\|X^1\| = \cdots = \|X^k\| = 1} \|\mathrm{D}^k \operatorname{per} A(X^1, \ldots, X^k)\|. \tag{2.20}$$

Using (2.14) and the facts if $\|X^j\| = 1$ for all j then $\|P_k(X^1 \vee \cdots \vee X^k)P_k\| \leq 1$ and norm of a submatrix of a matrix is less than or equal to the norm of the matrix, we get

$$\begin{aligned}
\|D^k \operatorname{per} A\| &= k! \sup_{\|X^1\| = \cdots = \|X^k\| = 1} |\operatorname{tr}[(P_{n-k}(\tilde{\vee}^{n-k} A) P_{n-k}) \\
&\qquad\qquad (P_k(X^1 \vee \cdots \vee X^k) P_k)]| \\
&\leq k! \, \|P_{n-k} \, (\tilde{\vee}^{n-k} A) \, P_{n-k}\|_1 \\
&\leq k! \binom{n}{k} \|P_{n-k} \, (\tilde{\vee}^{n-k} A) \, P_{n-k}\| && \text{(using (2.18))} \\
&\leq k! \binom{n}{k} \|\tilde{\vee}^{n-k} A\| \\
&= k! \binom{n}{k} \| \vee^{n-k} A\| \\
&= k! \binom{n}{k} s_1(A)^{n-k} && \text{(by (2.19))} \\
&= k! \binom{n}{k} \|A\|^{n-k}.
\end{aligned}$$

□

As a corollary, we obtain a perturbation bound for per.

Corollary 2.1. *Let A and X be $n \times n$ matrices. Then*

$$|\operatorname{per}(A+X) - \operatorname{per} A| \leq \sum_{k=1}^{n} \binom{n}{k} \|A\|^{n-k} \|X\|^{k}. \tag{2.21}$$

Proof. This follows using Taylor's theorem.

$$\begin{aligned} |\operatorname{per}(A+X) - \operatorname{per} A| &= \| \sum_{k=1}^{n} \frac{1}{k!} D^k \operatorname{per} A(X, \dots, X) \| \\ &\leq \sum_{k=1}^{n} \frac{1}{k!} \| D^k \operatorname{per} A(X, \dots, X) \| \\ &\leq \sum_{k=1}^{n} \binom{n}{k} \|A\|^{n-k} \|X\|^{k}. \end{aligned}$$

□

Consider the simplest commutative case: $A = I$, $X = xI$. Then the expression on both the sides of the inequality (2.21) is

$$\sum_{k=1}^{n} \binom{n}{k} x^k.$$

So no improvement of the corollary is possible in this sense.

3 Coefficients of the characteristic polynomial

The *characteristic polynomial* of A, by definition, is

$$x^n - g_1 x^{n-1} + g_2 x^{n-2} - \ldots + (-1)^n g_n, \tag{3.1}$$

where the coefficient g_r is the sum of $r \times r$ principal minors of A. In particular, g_1 is the trace of A and g_n is the determinant of A. We consider $g_r : \mathbb{M}(n) \to \mathbb{C}$ as the map taking a matrix to the r^{th} coefficient of its characteristic polynomial. Then

$$g_r(A) = \sum_{\mathcal{I} \in Q_{r,n}} \det A_{\mathcal{I}},$$

where $A_{\mathcal{I}}$ denotes the submatrix $A[\mathcal{I}|\mathcal{I}]$ of A.

For $\mathcal{I} = (i_1, i_2, \ldots, i_r) \in Q_{r,n}$, let $h_{\mathcal{I}}$ denote the map which takes an $n \times n$ matrix A to $A_{\mathcal{I}}$. It is a linear map. Then,

$$g_r(A) = \sum_{\mathcal{I} \in Q_{r,n}} (\det \circ h_{\mathcal{I}})(A). \tag{3.2}$$

We derive three different expressions for higher order derivatives of the coefficients of the characteristic polynomial of A, which follow as corollaries to the theorems in [7]. We first give a lemma by which all three of them follow immediately.

Lemma. *If f and g are two maps such that $f \circ g$ is well defined and g is linear, then*

$$\mathrm{D}^k(f \circ g)(A)(X^1, \ldots, X^k) = \mathrm{D}^k f(g(A))(g(X^1), \ldots, g(X^k)). \tag{3.3}$$

Proof. This follows by (1.2). □

From theorems 1 and 2 of [7], we obtain

Theorem 3.1. *For $1 \le k, r \le n$,*

$$\mathrm{D}^k g_r(A)(X^1, \ldots, X^k) = \sum_{\mathcal{I} \in Q_{r,n}} \sum_{\mathcal{J} \in Q_{k,r}} \sum_{\sigma \in S_k} \det(A_{\mathcal{I}})(\mathcal{J}; X_{\mathcal{I}}^{\sigma(1)}, \ldots, X_{\mathcal{I}}^{\sigma(k)}). \tag{3.4}$$

Theorem 3.2. *For $1 \le k, r \le n$,*

$$\mathrm{D}^k g_r(A)(X^1, \ldots, X^k) = \sum_{\mathcal{I} \in Q_{r,n}} \sum_{\mathcal{J}, \mathcal{K} \in Q_{k,r}} \sum_{\sigma \in S_k} (-1)^{|\mathcal{J}|+|\mathcal{K}|} \det(A_{\mathcal{I}})(\mathcal{K}|\mathcal{J}) \det(Y^{\sigma}_{[\mathcal{J}]})_{\mathcal{I}}[\mathcal{K}|\mathcal{J}]. \tag{3.5}$$

We now give another expression for the derivatives of g_r using theorem 3 of [7]. Let $\wedge^k \mathcal{H}$ denote the antisymmetric tensor power of $\mathcal{H}$. If $\{e_i\}$ is an orthonormal basis of $\mathcal{H}$, then for $\mathcal{I} = (i_1, \ldots, i_k) \in Q_{k,n}$, we define $e_{\mathcal{I}} = e_{i_1} \wedge \cdots \wedge e_{i_k}$. Then, $\{e_{\mathcal{I}} : \mathcal{I} \in Q_{k,n}\}$ is an orthonormal basis of $\wedge^k \mathcal{H}$. It is conventional to order these multi-indices lexicographically. (See [4], Chapter 2.) $\wedge^k A$ is the kth *antisymmetric tensor power* of A. With respect to the above mentioned basis, the $(\mathcal{I}, \mathcal{J})$-entry of $\wedge^k A$ is $\det A[\mathcal{I}|\mathcal{J}]$. The transpose of the matrix with entries $(-1)^{|\mathcal{I}|+|\mathcal{J}|} \det A(\mathcal{I}|\mathcal{J})$ can be identified with an operator on the space $\wedge^k \mathcal{H}$. We call this operator $\tilde{\wedge}^{n-k} A$ and note that it is unitarily similar to the transpose of $\wedge^{n-k} A$.

For $X^1, \ldots, X^k \in \mathscr{L}(\mathcal{H})$, consider the operator given by (2.13). This leaves the space $\wedge^k \mathcal{H}$ invariant. We use the notation $X^1 \wedge X^2 \wedge \cdots \wedge X^k$ for the restriction of the operator (2.13) to the subspace $\wedge^k \mathcal{H}$.

Theorem 3.3. *For $1 \le k, r \le n$,*

$$\mathrm{D}^k g_r(A)(X^1, \ldots, X^k) = k! \sum_{\mathcal{I} \in Q_{r,n}} \mathrm{tr}\, [(\tilde{\wedge}^{r-k} A_{\mathcal{I}})(X^1_{\mathcal{I}} \wedge \cdots \wedge X^k_{\mathcal{I}})]. \tag{3.6}$$

In particular,

$$\mathrm{D}^k g_r(A)(X, \ldots, X) = k! \sum_{\mathcal{I} \in Q_{r,n}} \mathrm{tr}\, (\tilde{\wedge}^{r-k} A_{\mathcal{I}})(\wedge^k X_{\mathcal{I}})].$$

Let $s_1(A) \geq \cdots \geq s_n(A)$ be the singular values of A, and let $\|A\| := s_1(A)$ be the *operator norm* of A. Also,

$$\|\mathrm{D}g_r(A)\| = \sup_{\|X\|=1} \|\mathrm{D}g_r(A)(X)\|. \tag{3.7}$$

The *trace norm* of A is defined as

$$\|A\|_1 = s_1(A) + \cdots + s_n(A). \tag{3.8}$$

This is the dual of the operator norm ([4], Chapter 4). So,

$$\|A\|_1 = \sup_{\|X\|=1} |\operatorname{tr} AX|.$$

For $1 \leq k \leq n$, let $p_k(x_1, \ldots, x_n)$ denote the kth elementary symmetric polynomial in n variables. We now give the exact norm of $\mathrm{D}^k g_r(A)$, using theorem 4 of [7].

Theorem 3.4. *Let A be an $n \times n$ matrix and let $s_1(A_{\mathcal{I}}), \ldots, s_r(A_{\mathcal{I}})$ be the singular values of $A_{\mathcal{I}}$. Then,*

$$\|\mathrm{D}^k g_r(A)\| = k! \sum_{\mathcal{I} \in Q_{r,n}} p_{r-k}(s_1(A_{\mathcal{I}}), \ldots, s_r(A_{\mathcal{I}})). \tag{3.9}$$

Proof. For every $\mathcal{I} \in Q_{r,n}$, let $A_{\mathcal{I}} = UA_{\mathcal{I}}^+ V$ be the singular value decomposition of $A_{\mathcal{I}}$. Using (1.2) and (3.2), we get

$$\mathrm{D}^k g_r(A)(X^1, \ldots, X^k) = \sum_{\mathcal{I} \in Q_{r,n}} \mathrm{D}^k \det A_{\mathcal{I}}^+ (U^* X_{\mathcal{I}}^1 V^*, \ldots, U^* X_{\mathcal{I}}^k V^*).$$

It follows from here that

$$\|\mathrm{D}^k g_r(A)\| = \|\mathrm{D}^k \det A_{\mathcal{I}}^+\|.$$

Theorem 4 of [7] now gives the desired result. ☐

As a corollary we have the following perturbation bound for g_r.

Corollary 3.1. *Let A and X be $n \times n$ matrices. Then,*

$$|g_r(A+X) - g_r(A)| \leq \sum_{\mathcal{I} \in Q_{r,n}} \sum_{k=1}^{r} p_{r-k}(s_1(A_{\mathcal{I}}), \ldots, s_r(A_{\mathcal{I}})) \|X\|^k. \tag{3.10}$$

Proof. This is again a consequence of Taylor's theorem.

$$g_r(A+X) = g_r(A) + \sum_{k=1}^{r} \frac{1}{k!} \mathrm{D}^k g_r(A)(X, \ldots, X) + O(||X||^{r+1}). \tag{3.11}$$

It follows from here that

$$|g_r(A+X) - g_r(A)| \leq \sum_{k=1}^{r} \frac{1}{k!} \|\mathrm{D}^k g_r(A)\| \|X\|^k.$$

Using theorem 3.4 in this, we get (3.10). ☐

In the simplest commutative case where $A = I$ and $X = xI$, both sides of (3.10) equal

$$\sum_{k=1}^{r} \binom{n}{r}^2 x^k.$$

A weaker perturbation bound can be obtained as follows.

Corollary 3.2. *For $n \times n$ matrices A and X,*

$$|g_r(A+X) - g_r(A)| \leq \sum_{k=1}^{r} \binom{n}{r} \|A\|^{r-k} \|X\|^k. \tag{3.12}$$

Proof. This follows by corollary 2 and by using the facts that $\|A\| = s_1(A)$ and norm of a submatrix of a matrix is less than or equal to the norm of the matrix. □

Acknowledgement

The possibility of studying these functions on the same lines as determinants was suggested by Rajendra Bhatia. It is a pleasure to record my thanks to him.

References

[1] R. Bhatia, *Matrix Analysis*, Springer, New York, 1997.

[2] R. Bhatia, *Positive Definite Matrices*, Princeton University Press, New Jersey, 2007.

[3] R. Bhatia, *Perturbation Bounds for Matrix Eigenvalues*, SIAM, Philadelphia, 2007, expanded reprint of 1987 edition.

[4] R. Bhatia, Variation of symmetric tensor powers and permanents, *Linear Algebra and its Applications* 62(1984)269-276.

[5] R. Bhatia and J.A. Dias da Silva, Variation of induced linear operators, *Linear Algebra and its Applications* 341(2002)391-402.

[6] R. Bhatia and S. Friedland, Variation of Grassman powers and spectra, *Linear Algebra and its Applications* 40(1981)1-18.

[7] R. Bhatia and T. Jain, Higher order derivatives and perturbation bounds for determinants, *Linear Algebra and its Applications* 431(2009)2102-2108.

[8] J. Dieudonné, *Foundations of Modern Analysis*, Academic Press, New York, 1960.

[9] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities*, Dover Publications, New York, 1992, reprint of 1964 edition.

[10] R. Merris, *Multilinear Algebra*, Gordon and Breach Science Publishers, Singapore, 1997.

[11] H. Minc, *Permanents*, Addison-Wesley Publishing Company, Massachusetts, 1970.